Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.583 – 598

RESEARCH ARTICLE

EXPERT SYSTEM FOR CORONARY HEART DISEASE - BUILT USING ARTIFICIAL INTELLIGENCE

Suchithra[1], Dr. P.Uma Maheswari[2]

[1]PG Scholar, Info Institute of Engineering, India

[2]Professor and Head, Department of CSE, Info Institute of Engineering, India

chanu.suchithra@gmail.com

dr.umasundar@gmail.com

Abstract — Coronary Heart Disease is a disease which is difficult to diagnose and is vey commonly identified only during the mortality of an individual. The World Health Organization (WHO) reported that 70 per cent deaths occur in subjects less than 70 years of age in India and in other developing countries. Since Coronary Heart Disease (CHD) is becoming an epidemic in India, there is a terrific need for effective solution for risk identification as earlier as possible. Despite ccomputerized clinical guidelines may provide benefits for health outcomes and costs, however, their successful implementation are more challenging to investigate significant problems. One effective solution is to achieve an optimal trade-off between data ambiguity and good decision-making which would further help in the integration of data mining and artificial intelligence techniques. In this work, a novel approach is proposed to develop a clinical decision support system (CDSS) for heart disease diagnosis using data mining and Artificial Intelligence techniques.The major goal of this paper is to build an expert system for diagnosing the presence of Ischemic Heart Disease with an integrated automated classifier using Artificial Intelligence techniques. A retrospective data set that included 1000 clinical cases is taken for the work. 80 sets were discarded during preprocessing. Tests were run on 920 cases using weka classifiers [5] available in weka 3.7.0. The proposed algorithm formalizes the treatment of vagueness in decision support architecture which also evaluates the performance measures among them. Experimental results demonstrate the effectiveness of the proposed CDSS in heart disease diagnosis.

Keywords— Artificial Intelligence Techniques, Clinical Decision Support System (CDSS), Coronary Heart Disease

I. INTRODUCTION

India at present is under at a transition which is most epidemiological and one among this is the epidemic cardiovascular disease. The statistical and probabilistic cause-specific mortality rate data indicate that cardiovascular disease plays an important role in contribution to mortality. Cardiovascular Diseases (CVD) includes a group of diseases of the heart and

vascular system. Indians are more likely to have these types of heart disease that would lead to worst outcomes like ischemic heart disease — a condition which is characterized by reduced blood supply to heart.

Coronary Heart Disease (CHD) or Ischemic Heart Disease (IHD) causes 25-30 percent of deaths in most industrialized countries. India is in a risk of developing more deaths due to CHD. Recent Studies say that nowadays sudden death occur mainly during sleep which has become prevalent due to the lack of oxygen supply to the heart.

Therefore, a Decision Support System (DSS) is being proposed for the identification of the level of risk in Ischemic Heart Disease for a Patient as much earlier as possible. This would help the patients in taking precautionary steps like following a balanced diet and medications which will in turn may increase the life time of a patient.

This work focuses on different classification algorithms which are applied on a dataset that are collected from Coimbatore Medical College. A comparison of different measures like sensitivity, specificity kappa statistic, ROC[9], time taken which are taken for classification is compared with different classification algorithms.

.

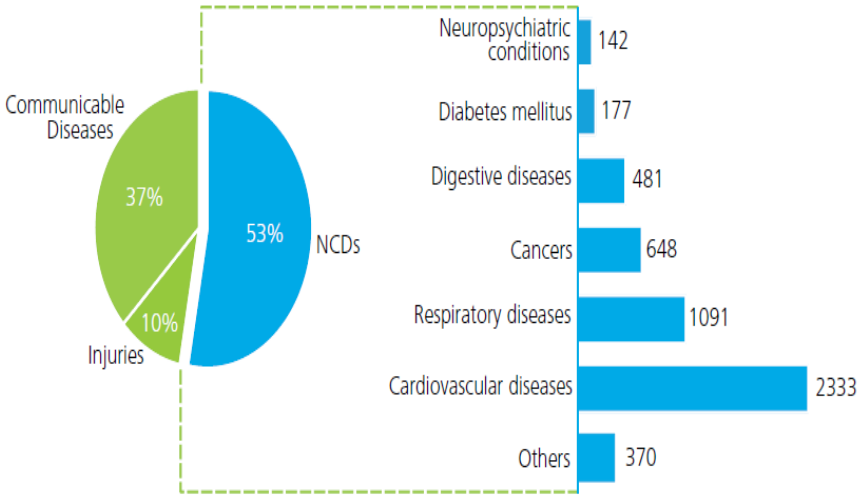


Fig 1.1 Causes of death in India

This work undergoes the classification of risk level for heart disease. Knowing about the risk level, people can take the adequate measures that are required for diagnosing heart disease or following very strict dietary rules to avoid future problems that might cause morality.

Artificial Intelligence [1][3][5][6] is particularly useful in medicine when there are no dispositive evidence favouring a particular treatment option. Based on patients' profile, history, physical examination, diagnosis and utilizing previous treatment patterns, new treatment plans can be effectively suggested.

Artificial Intelligence is a collection of algorithmic ways to extract informative patterns from raw data.

Artificial Intelligence is purely data-driven; this feature is more important in health care.

Artificial Intelligence has a set of tools and techniques that are applied for processing data for discovering hidden patterns that will provide healthcare professionals an additional source of knowledge for decision making[10].

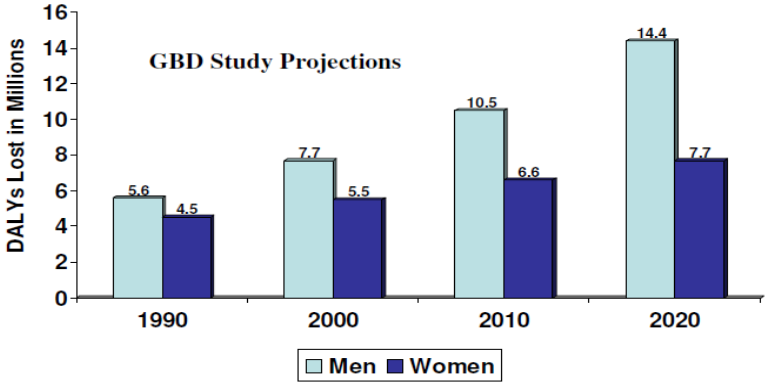


Fig 1.2 Increasing CVDs in India

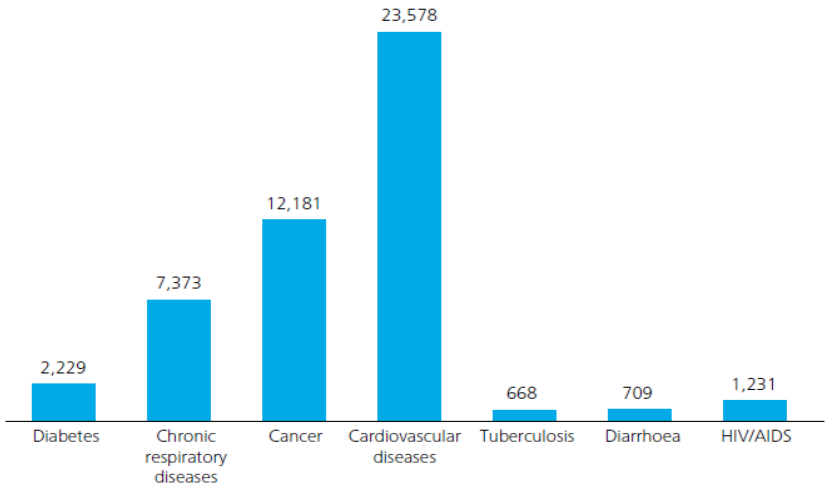


Fig 1.3 Morality of different Diseases

II. RELATED WORK

The following have been analysed and studied in order to develop an expert system for heart disease in India.

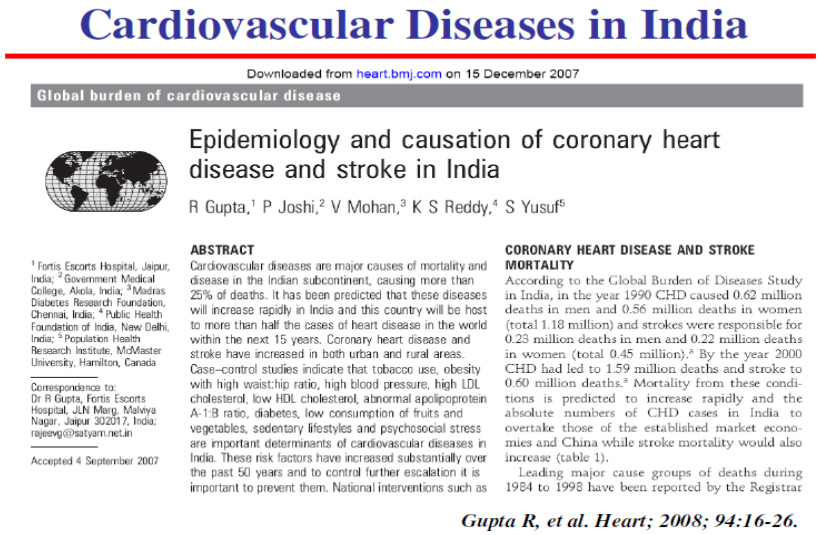
The World Health Statistics as released in 2012 by the World Health Organization (WHO) clearly enumerate the worrying trends of Indians. It is said that among adults:

- Over 20 years of age - the estimated presence of CHD is: 3-4% in rural areas & 8 -10% in urban areas, which represents a two-fold rise in rural areas and a six-fold rise in urban areas between the years of 1960 and 2000.

When compared with the population of United States, the presence of CHD in Asian Indians is approximately 4 times higher. Indian situation according to WHO statistics is : 24% men and 22.6% women - aged 25 years and above - are suffering from high blood pressure(BP).

As per Statistical Reports:

- One in 10 men and women aged 25 years and above has high sugar.
- More women in India (2.5% of adults aged 20 years and above) are obese compared to Indian men (1.3%).
- Almost one in five (19%) boys, aged 13-15 years (adolescents), and 8% girls smoke tobacco.



Cardiovascular Diseases in India

Downloaded from heart.bmj.com on 15 December 2007

Global burden of cardiovascular disease

Epidemiology and causation of coronary heart disease and stroke in India

R Gupta,[1] P Joshi,[2] V Mohan,[3] K S Reddy,[4] S Yusuf[5]

[1] Fortis Escorts Hospital, Jaipur, India; [2] Government Medical College, Akola, India; [3] Madras Diabetes Research Foundation, Chennai, India; [4] Public Health Foundation of India, New Delhi, India; [5] Population Health Research Institute, McMaster University, Hamilton, Canada

Correspondence to: Dr R Gupta, Fortis Escorts Hospital, JLN Marg, Malviya Nagar, Jaipur 302017, India; rajeevg@satyam.net.in

Accepted 4 September 2007

ABSTRACT

Cardiovascular diseases are major causes of mortality and disease in the Indian subcontinent, causing more than 25% of deaths. It has been predicted that these diseases will increase rapidly in India and this country will be host to more than half the cases of heart disease in the world within the next 15 years. Coronary heart disease and stroke have increased in both urban and rural areas. Case–control studies indicate that tobacco use, obesity with high waist:hip ratio, high blood pressure, high LDL cholesterol, low HDL cholesterol, abnormal apolipoprotein A-1:B ratio, diabetes, low consumption of fruits and vegetables, sedentary lifestyles and psychosocial stress are important determinants of cardiovascular diseases in India. These risk factors have increased substantially over the past 50 years and to control further escalation it is important to prevent them. National interventions such as

CORONARY HEART DISEASE AND STROKE MORTALITY

According to the Global Burden of Diseases Study in India, in the year 1990 CHD caused 0.62 million deaths in men and 0.56 million deaths in women (total 1.18 million) and strokes were responsible for 0.23 million deaths in men and 0.22 million deaths in women (total 0.45 million).[3] By the year 2000 CHD had led to 1.59 million deaths and stroke to 0.60 million deaths.[3] Mortality from these conditions is predicted to increase rapidly and the absolute numbers of CHD cases in India to overtake those of the established market economies and China while stroke mortality would also increase (table 1).

Leading major cause groups of deaths during 1984 to 1998 have been reported by the Registrar

Gupta R, et al. Heart; 2008; 94:16-26.

Fig 2.1 A wide spread CVD in India- Article

European Union experts declare that there still a drastic change in the disparities between the countries - not only in terms of CVD incidence, but also in regard to the national prevention policies. According to them, obese adolescents do not have symptoms of heart disease or any damaged hearts with thicker walls. Moreover, both structural and functional measures correlate with Body Mass Index (BMI). That is why these findings have an explicate reason why obesity is a risk for heart disease. Therefore, a long-term case fatality following acute coronary syndrome is considerably high among Indians when compared with the other country populations. In addition to this, there are several reversal of socio-economic gradients for the CHD factors that emerge in the Indian Population.

The problem that persists in the present situation is that less expert doctors play the major impact for the motivation of our problem statement.

The Statistical Report indicates the following:

- As per the Medical Council of India, the total number of registered allopathic doctors in the country is under 7 lakh and the current population of India is 1.22 Billion.
- About 72.2% of the population live in some 638,000 villages and the rest 27.8% in about 5,480 towns and urban agglomerations.
- According to this statistic, only one allopathic doctor is available for over 1,600 patients.
- As per the medical council the total allopathic doctors produced per year are 12,000.
- Therefore, this brings a miserable situation of measure as just 1 doctor for a population of 2,000.
- And the population of nurses, is just 1 for every 2,200 patients.

Dr.Vaidyanathan and K.Rajeswari et al, [1] 2012 stated in their work as: A heart system inquiry diagnosis scale is designed, where the symptoms are defined clearly, and therefore a detailed collecting methods are listed which helps in the identification of the CVD. The complete dataset is divided into training and testing data. A 10-fold cross validation is used here. Different classifiers in Weka are trained and tested with the training and testing datasets and therefore the sensitivity and the specificity of each classifier is evaluated using Ratio Analysis. This method help in reinforcing the validation process, such that the sensitivity and specificity values are with the average of ten validation folds. A training set is a set of data that is used for discovering potentially related predictive relationship between the given input and the produced output.

KS Reddy et al [2] submitted a statistics in 2011 as: The suggestion for the need of Information Technology and its implementation towards the medical field. This paper, specifies the need for a Decision Support System which is indeed in order to control the widespread epidemic CardioVascular Disease. The system should be designed for the Indian Population. The suggestion of this paper is that Coronary Heart Disease or Ischemic Heart Disease can be handled successfully if more researches are encouraged in this mixed medical cum technology area.

Dr.P.Amirtaraj et al [3] 2011 discussed Cardiovascular Diseases (CVD) contains of a group of diseases that concerns with the heart and vascular system. The major condition of this disease includes Coronary Heart Disease (CHD) or Ischemic Heart Disease(IHD) causes 25-30 percent of deaths (major part in death percentage) in most of the industrialized countries. India too is in a risk of developing more deaths due to CHD. Therefore a Decision Support System (DSS)[11] is being proposed in order for the identification of the level of risk of Ischemic Heart Disease for a Patient. This would help the patients in taking the required precautionary steps like: following a balanced diet, medication etc., which in turn may increase the life time of a patient. The attributes for predictions are being selected under certain factors i.e., after considering Indian conditions from literature and on the Expert advice from Doctors. Framingham Risk score is being used in this paper which includes five attributes for comparison. The system proposed here includes fourteen features which are to be analyzed according to Indian Conditions. The system is only a theoretical study which proposes implementation of Artificial Intelligence to mine the knowledge from the Medical data. This paper concludes that by the end of next year, India will have 60% of the world's heart disease burden. The statistical report says that when compared to people in other developed countries, the average age of patients with heart disease is at least 5-8 years lower among Indians than from the western areas. Normally, sixty is the average age for the heart patients in India against 63-68 in developed countries. In this scenario the age is slipping further to the mid-50s. Indians are more likely to have different types of heart disease like ischemic heart disease — a condition characterized by reduced blood supply to the heart.

Dr.P.Amirtharaj et al [4] 2012 concentrated on the following: Machine Intelligence is the area used in Medical Data Mining. Medical Data Mining term defines in finding interesting information from large collection of Medical data. This paper helps us to find useful information from heart disease data set. This paper specifies the theoretical study of the implementation of Machine Intelligence algorithms. This paper also states that the system proposed includes nineteen features. Therefore, for reducing the number of features Genetic Algorithm is being used. The importance of data mining techniques is to play an important role for providing better patient care and for an effective diagnostic capability by finding exact patterns and by extracting knowledge which would increase with the increase in the volume of stored data .

WHO 2010 [5] addresses the following: Coronary risk factors are widespread and this disease needs an urgent action for preventing a further rise in socioeconomic development proceeds. A Clinical Decision Support System (CDSS) for heart disease

which classifies the risk using Artificial Intelligent techniques have been proposed. This approach focuses on CAD Risk analysis, for a sample population, future work is directed for further analysis.

National Status:

- Until recent times, an ad-hoc disease survey, is being conducted which is a systematic approach and which is used for tracking disease does not exist.
- According to the journal publication in IJMR, it's been concluded that a national level critical components standard disease management guidelines, diet and physical activity guidelines are still not developed with intelligent computational assistance.

International Status:

- Substantial research work is in progress all over the world especially in countries like USA, UK, Europe, Japan and China. Since, there are no successful implementations still published, the products are much expensive and unaffordable by the public of India like countries.

By Rajeev Gupta & his associates [6] 2007 the following issues and its impacts were discussed: The population of India is more than 30 million people who are with diabetes. The number of diabetics climb steeply day by day. It is evaluated that by 2025, India will be in number one position in the world with the maximum number of people with diabetes. This estimation is in accordance to the World Health Organization (WHO)'s press release in 2012. According to the publication by WHO, people who are affected with diabetes in future will be in the age group 40 and 64. With the widespread of fast-food outlets and more sedentary lifestyles, the prevalence of diabetes in India is rising in hike. Therefore, a technological support is needed to prevent the on-growth of the life demanding Coronary Heart Disease.

T. Gurumoorthy et al [7] 2012 found a solution as: Diabetes Mellitus, is a metabolic disorder which is described by chronic hyperglycemia is a disease enhancing rapidly. It is a polygenic disease which is characterized by an abnormal high glucose in the blood. Statistics upgrade that 90 to 95 % of the World Diabetics have Type 2 Diabetes. And India tops the list above all countries in the statistical analysis. The consequences of diabetes is the macro vascular complication or micro vascular complication. A new model for the prediction of complications which are developed due to Diabetes Mellitus is proposed. Data collection is done after discussion with the Diabetologists of Diabetes care and Research centre. A questionnaire pattern is prepared with seven stages which after getting concern from Expert Doctor's opinion and Artificial Neural Network technique is used to predict the complications that are developed. Many researchers have contributed for the diagnosis of Type 2 Diabetes, this work focuses on modelling an effective Diagnosis which is of a special complication called neuropathy.

Dr.V.Vaithiyanathan and Dr.P.Amirtharaj et al [8] 2012 found the following: A Decision Support System (DSS) is proposed which is highly needed for the identification of the level of risk in Ischemic Heart Disease for a Patient. Framingham Risk score involves the usage of five attributes which is used for comparison. The proposed system consists of seventeen features which is analyzed accordingly to the Indian Conditional factors. This paper proposes the implementation of Artificial Neural Network technique for mining the knowledge from the collected Medical data.

. K. Rajeswari & V. Vaithiyanathan [9] 2010 et al stated the following as: Clinical guidelines which are computerised usually provide the most effective implementation includes much more significant problems. In this paper, a clinical decision support system (CDSS) is proposed for effective diagnosis of HD using Data Mining and AI techniques. The proposed CDSS utilizes Association pattern mining, GA and fuzzy logic (FL) for the effective diagnosis of the disease. The input to the system is a clinical dataset which contains the medical records of heart patients. Primarily, the dataset is being preprocessed for the elimination of noisy and insignificant information, such as the null values, irrelevant data (i.e., data entry errors, out of range data) and more from the data set, which is prepare for the adequate mining process. The two major criterions that are identified which affect the medical diagnosis are vagueness and uncertainty of medical records. In order to fix these problems, we employ Apriori, a standard association rule mining (ARM) algorithm, which are used to discover association patterns that are significant to HD diagnosis. Subsequently, these significant patterns are provided as initial population to GA. The GA produces a set of high impact parameters and their optimal values which serve as fuzzy inputs. Then, for each fuzzy input, a corresponding membership function and the fuzzy rule sets are generated. Finally, a patient data record is provided as input to the Fuzzy system, it deduces the risk of the HD.

The key objective presented by Preethi S.J. and K.Rajeswari [10] 2009 is for medical image processing in such a way that the results obtained will be more suitable than the original image which is given for a medical application, that can also be achieved only by applying range compression, contrast stretching, histogram equalization, noise smoothing algorithms. Digital Medical images[22] which are usually affected by unwanted noise, blurriness which also suffer from lack of contrast and sharpness; sometimes results in false diagnosis. This paper concentrates mainly on making diagnosis easy by eliminating the above specified problems.

III.MATERIALS AND METHODS

The development of Clinical Decision Support System for Ischemic Heart Disease (IHD) diagnosis consists of four phases: phase I, phase II, phase III and phase IV. As in Fig 3.1 Phase I, includes the collection of an attribute set for IHD.

Phase II, data is being collected ; Training and testing databases were constructed. In Phase III, the data is being coded. A training database is being tested with a set of automated classifiers. In Phase IV, the developed Decision Support System is being evaluated. In this paper, a heart system inquiry diagnosis scale [5] is being designed, where the symptoms are clearly defined, and the collection methods are listed in detail. The total datasets are divided into two: training and testing data. A 10-fold cross validation is used. The different classifiers under Weka are trained using the training dataset and the sensitivity and the specificity of the classifiers are evaluated on the test set. This evaluation is repeated ten times, alternating the test set used each time. This method reinforces the validation process, so that the sensitivity and specificity values are the average of the ten validation folds [21].A training set is a set of data used to discover potentially predictive relationships [22] between the input and output. The training set is used by the classifiers of Weka to build a model. This model created is being used to predict the output for the test attributes.

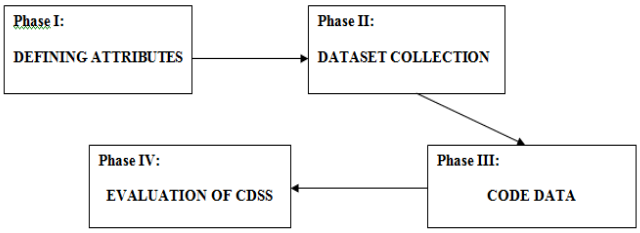


Fig: 3.1 Phases in Development of CDSS

3.1 Phase I – Defining Attributes

The First stage of the work is attribute identification. Various literatures and bench mark UCI datasets [23] were analyzed and discussed with experts. However, all the published experiments refer to the 13 attributes of these, as in Table 3.1. The identified attributes will help in designing clinical decision support system demonstrated in Table 3.2. These attributes can be easily collected through medical camps in a particular area. The method is feasible under the assumption that already people are familiar with the presence of cholesterol and presence of diabetes. In future, these measures as values also can be obtained through instruments available in market.

No	Name	Description
1	Age	age in years
2	Sex	1 = male ; 0 = female
3	Cp	chest pain type (1 = typical angina; 2 = atypical angina ; 3 = non-anginal pain; 4 = asymptomatic)
4	Trestbps	resting blood pressure (in mm Hg on admission to the hospital)
5	Chol	serum cholesterol in mg dl
6	Fbs	(fasting blood sugar > 120 mg dl) (1 = true; 0 = false)
7	Resteog	resting electrocardiographic results (0 = normal; 1 = having ST-T wave abnormality; 2 = showing probable or define left ventricular hypertrophy by Estes' criteria)
8	Thalach	maximum heart rate achieved
9	Exang	exercise induced angina (1 = yes; 0 = no)
10	OldPeak	ST depression induced by exercise relative to rest
11	Slope	the slope of the peak exercise ST segment (1 = upsloping; 2 = flat ; 3= downsloping)
12	Ca	number of major vessels (0-3) colored by fluoroscopy
13	Thal	(3 = normal; 6 = fixed defect; 7 = reversible defect)
14	Num	(3 = normal; 6 = fixed defect; 7 = reversible defect)

Table 3.1 – Attribute Identification

3.2 Phase II – Dataset Collection

A retrospective database (DB1) comprising the records of 920 clinical cases was recorded for algorithm training. Confirmed IHD cases accounted for 68% of all records, while the remaining other diagnoses had signs, symptoms and complaints that were shared with clinical cases of IHD. A testing database(DB2) was created for classifier evaluation.

No	Name	Description
1	Age	age in years
2	Sex	1 = male ; 0 = female
3	Menopause	chest pain type (1 = typical angina; 2 = atypical angina ; 3 = non-anginal pain; 4 = asymptomatic)
4	Height	resting blood pressure (in mm Hg on admission to the hospital)
5	Weight	serum cholesterol in mg/dl
6	BMI	Body Mass Index
7	Waistcircum	Measure of circumference of Waist
8	SBP	Systolic Blood Pressure
9	DBP	Diastolic Blood Pressure
10	Diabetes	Presence is treated as 1, absence as 0.
11	Chol	Presence of cholesterol is treated as 1, absence as 0.
12	Thy	Presence is thyroid is treated as 1, absence as 0.
13	Per_hab	Personal habits of drinking/smoking is recorded as 1, 0 otherwise
14	Fam_hist	Family history of heart disease presence is treated as 1, 0 otherwise.
15	Type A	Type A personality or person with sleeping disorder is treated as 1, 0 otherwise.
16	Out	Output, No risk as 0, Low risk as 1, Medium risk as 2 and High risk as 3. Later for 2 class problem, No risk as 0 and the others as 1

Table 3.2 – Attribute Description

An experimental model is being designed for the determination of the classifiers that are to be integrated into the system of the CDSS which is under development. Automated classifiers available from the Waikato Environment for Knowledge Analysis (Weka 3.7.0) software [14] are tested with the dataset. A total of 16 attributes (nominal, numeric) of signs, symptoms and high-risk groups which are listed, coded and given ―y(yes) or ―n (no) values depending on whether the patient had a certain symptom. Certain attributes like height, weight, body mass index, systolic and diastolic blood pressure are numeric values. The target variable was the diagnostic variable out, with Risk or No Risk categories.

3.3 Phase III - Training and Testing Classifiers

3.3.1 Artificial Intelligence Techniques

- Bayesian classifiers: Classification is being achieved through the calculation of the probability for each class, assuming conditional independence of the attributes in the NaiveBayes, NaiveBayesSimple and BayesNet algorithms [19]. A total of seven different Bayes classifier models that are available in weka is being tested for various measures like sensitivity, specificity, accuracy, kappa statistic as given in Table 3.2.
- Function classifier: Function classifiers include four classifiers namely multilayer perceptron, logistic regression, support vector machines and regression technique .Neural Networks are used for both classification and prediction. In this study, the parameters of learning rate (L), varying between 0.3 and 0.5, momentum (M), varying between 0.2 and 0.5, and 500 epochs to training (default), were applied to the algorithm Multilayer Perceptron (MLP).
- Lazy classifiers: Lazy classifiers use distance measure for find the training instance closest to the given test instance and predicts the same class as the training instance. They are instance based learners. Probability measures are also used. A total of 5 lazy classifiers are available in weka as shown in Table 3.3. IBk[15], LWL, KSTAR, LBR[14],J48 classifiers are used and J48 has demonstrated better accuracy for our data set.

Classifier Name
J48
KSTAR
IBK
IBI
Random Tree
Id3
Random Commit Tree
Rules
Random Forest
Nnge
Decorate
Classification via Regression
FT
NB
Rotation Forest
LMT
MLP

Table 3.3 – Different Classifiers used for Classification

- Meta classifiers: A total of 17 classifier models are available in weka. Meta classifiers are used to solve the problem of reliability with base classifiers. ―metalearning schemes enable users to combine instances of one or more of the basic algorithms in various ways: bagging, boosting and stacking. Vote meta classifier is a classifier combining classifiers using unweighted average of probability estimates or numeric predictions. A method called ―Filtered Classifier allows a filter to be paired up with a classifier. Classification can be made cost-sensitive, or multi-class, or ordinal-class. Parameter values which are selected can be used for cross-validation.
- Tree classifiers: A total of 14 classifiers are available in weka. A well accepted method of classification is the induction of decision trees [16, 17]. A decision tree is a flow chart like structure consisting of internal nodes, leaf nodes, and branches. Each internal node represents a decision, or test, on a data attribute, and each outgoing branch corresponds to a possible outcome of the test. Each leaf node represents a class. In order to classify an unlabeled data sample, the classifier tests the attribute values of the sample against the decision tree. A path is traced from the root to a leaf node which holds the class predication for that sample. Decision trees can easily be converted into IFTHEN rules [18] and used for decision making.
- Rules classifiers: A total of 10 models are available for classification in weka. A dataset which consists in couples x and y, where x is each element of the population and y the class it belongs to, a classification rule can be considered as a function that assigns its class to each element. A binary classification is such that the label y can take only a two values. A classification rule or classifier is a function h that can be evaluated for any possible value of x, specifically, given the data , h(x) will yields a similar classification as close as possible to the true group label y. The true labels yi can be known but will not necessarily match their approximations . In a binary classification, the elements that are not correctly classified are named false positives and false negatives.

ssifier	Correctly cl.	Incorrectly cl	Kappa	Precision	Recall	F-measure	ROC	TP	FP	FN	TN	Sensitivity	Specificity	Accuracy	Precision	Time
IBK	0.972222	0.027778	0.9426	0.972	0.972	0.972	0.991	523	11	17	369	0.97157191	0.97317073	0.9722222	0.981735	0 seconds
KSTAR	0.973214	0.026786	0.9448	0.973	0.973	0.973	0.995	545	14	13	348	0.9800338	0.96942686	0.9732143	0.977419	0 seconds
J48	0.9835233	0.0164767	0.9501	0.975	0.975	0.98	0.997	579	16	15	310	0.98447801	0.97046521	0.9835233	0.983351	0 seconds
IB1	0.962254	0.037746	0.9345	0.968	0.968	0.968	0.967	518	16	16	370	0.97517297	0.96153846	0.962254	0.975297	0 seconds
LBR	0.9350791	0.0649209	0.8698	0.937	0.937	0.937	0.972	494	40	24	362	0.95833333	0.90940741	0.9350791	0.93243	.05 seconds
LWL	0.9156746	0.0843254	0.8289	0.921	0.916	0.916	0.968	468	66	19	367	0.96513761	0.8594514	0.9156746	0.88851	0 seconds

Table 3.4 – Lazy Classifiers

3.3.2. Algorithm selection

The algorithm with the greatest area under the receiver operating characteristic curve (AUC) is selected. The receiver operating characteristic curve is a graphical plot of sensitivity versus one minus specificity, and is used to evaluate classification and prediction models [16]. In the Weka 3.7.0 software, the AUC is calculated by means of the sensitivity and specificity of each of the folds is used in training the classifier. The CDSS was designed such that it would issue a warning when the probability of IHD occurrence was higher than 50%.

3.4. Phase IV – Evaluation of CDSS

The CDSS was then evaluated using the testing dataset DB2. The data relating to each clinical case in DB2 were recorded in the CDSS. The probabilities of IHD occurrence were recorded for comparison against the diagnostic intuition that had been recorded by a medical provider during a consultation.

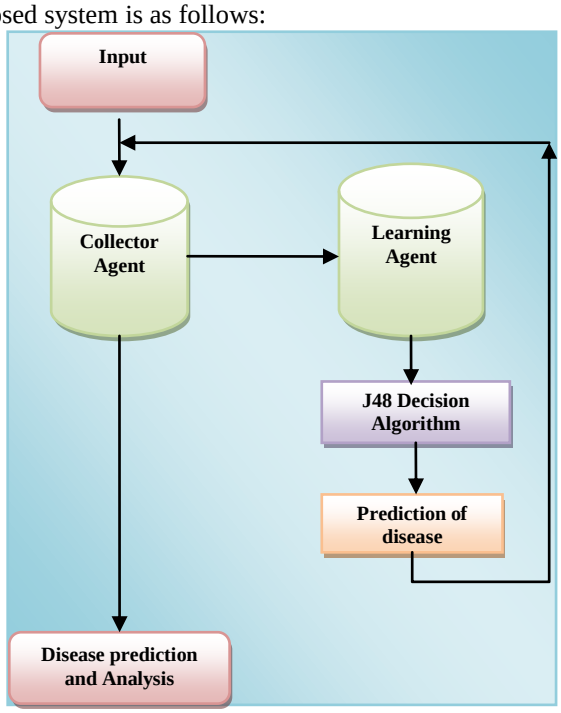
3. 5 Statistical analysis

Cross-validation with 10 subsets (10-fold cross-validation) is being used for testing. The evaluation and comparison of algorithms regarding the accuracy of the classification, the parameters used for selecting the algorithm were of the highest values for the AUC, sensitivity, specificity and accuracy rate [24]. Comparative analysis between the AUC results for each algorithm before and after variable selection is being performed for the determination of the best parameters for the algorithm to be selected. Kappa statistics were used to evaluate the agreements between the CDSS and the gold standard; between the physician impression, i.e. relating to the diagnostic impression recorded by physicians and their tutors during consultations, and the gold standard; and between the CDSS and the physician impression. In this manner, the diagnostic reliability of the CDSS is being assessed.

IV. PROPOSED SYSTEM

The data warehouse of heart disease includes the screening data of heart disorder patients who are under evaluation. Initially, this data warehouse is being pre-processed either manually or using data mining tool such as Rapid Miner to make the mining process more efficient. The algorithms are applied directly to the dataset. The data are being classified using the data mining tool. Usually WEKA data mining tool is widely used for classification. The dataset is generally split into training and test data. 10-fold cross validation is used in this work, where the process is out of 10 tuples, 9 are taken as training samples and 1 as testing sample. Evaluation is usually described in terms of accuracy. The main goal of different classification algorithms is accuracy improvement. As medical knowledge is vague, fuzzy sets are mainly used to deal with the uncertain linguistic medical concepts such as Less, Very Less, Medium and High.

The Overall Architecture of the proposed system is as follows:



Advantages:

- Helps in earlier identification of heart diseases.
- Reduction of death rate due to cardiovascular disease.

V. RESULTS AND DISCUSSIONS

A total of 59 models are used for classification of Ischemic disease and 16 models are found to be more efficient with accuracy when compared with physician‘s diagnostic impression and gold standard, kappa statistic (minimum of 0.736). J48 algorithm showed the best diagnoses with the highest accuracy 98.35%, sensitivity 0.984, specificity 0.97 kappa 0.952 and ROC 0.997. Therefore, its concluded that a clinical decision support system (CDSS) can be developed for assisting the expert physicians to separate the positive and negative cases of heart disease. There was sensible agreement between the physician‘s diagnostic impression and CDSS k = 0.45 (p = 0.0008).Sensitivity, specificity and accuracy are the commonly used statistical measures to illustrate the medical diagnostic test for enumerating how the test was good and consistent.

The Metrics are Calculated as follows:
Sensitivity= TP/(TP+FN);
Specificity = TN/ (TN+FP);
Accuracy= (TN+TP)/ (TN+TP+FN+FP)
Precision = TP/(TP+FP);
Recall = TP/(TP+FN);

where TP, TN, FN and FP denotes True Positive, True Negative, False Negative and False Positive respectively.

A comparison of 16 different best classifiers based on their performance measures such as sensitivity, specificity and accuracy are shown in Table 5.1

Sl.No.	Classifier Name	Correctly Classified	Incorrectly Classified	Sensitivity	Accuracy	Precision
1	J48	98.14	1.86	98	97.32	98
2	KSTAR	97.32	2.68	97.8	97.32	97.77
3	IBK	97.22	2.78	97.16	97.22	98.14
4	IBI	96.83	3.17	97.3	96.83	97.3
5	Random Tree	97.123	2.877	97.15	97.12	97.97
6	Id3	97.22	2.68	97.47	97.32	97.97
7	Random Commit Tree	97.22	2.78	97.16	97.22	98.14
8	Rules	96.92	2.88	95.62	97.12	96.66
9	Random Forest	97.02	2.98	97.15	97.02	98.2
10	Nnge	97.02	2.98	97.15	97.02	97.8
11	Decorate	95.83	4.17	96.93	95.83	95.95
12	Classification via Regression	95.54	4.46	96.28	95.54	96.11
13	FT	95.83	4.17	96.93	95.83	95.95
14	NB	95.83	4.17	96.93	95.83	95.95
15	Rotation Forest	95.83	4.17	96.45	95.83	95.45
16	LMT	96.73	3.27	97.13	96.73	97.3
17	MLP	97.02	2.98	97.31	97.02	97.64

Table 5.1- Comparison of 17 best Classifiers

Five AI techniques in classification (Bayes classifier, Function classifier, Meta classifier, Rule classifier and Tree classifier) were used with a total of 59 models. The algorithms were analyzed & were run and parameters are obtained for further analysis. WAODE classifier gives the best accuracy of 0.931 than Bayes Classifier. Fig. 5.1 shows the comparison of different algorithms on the basis of measures like sensitivity, specificity, accuracy and precision. WAODE algorithmic model proves to be the best among the 7 algorithms and Bayes Classifier with an accuracy of 93%.

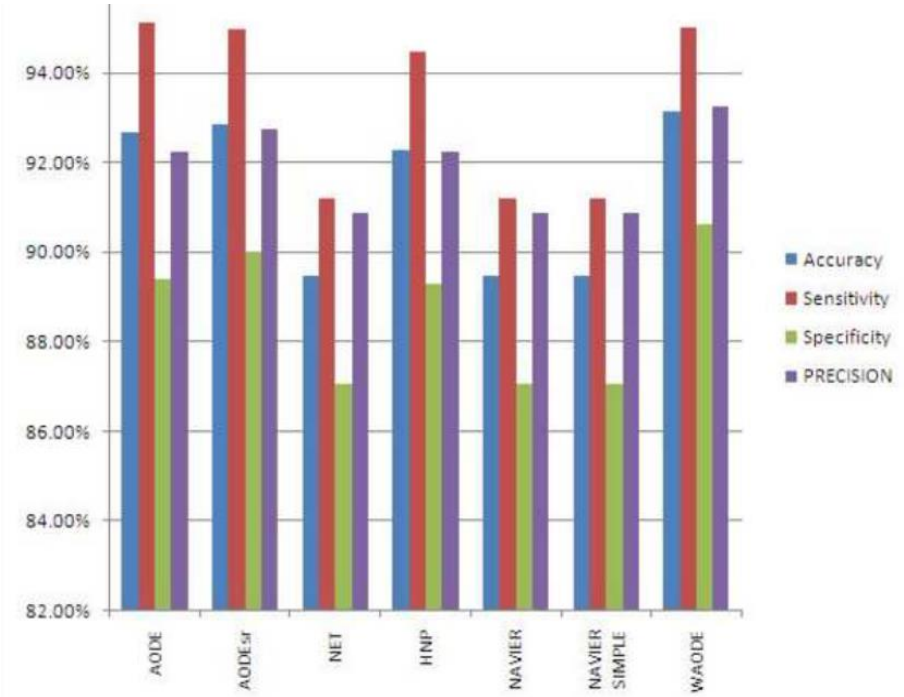


Fig 5.1 – Comparison of different Classifiers.

Fig. 5.2 Shows the comparison of various measures between different lazy classifiers. KSTAR and IBK are equally better than the other 3 models with an accuracy of 97%. J48 provides an accuracy of 98.35%.

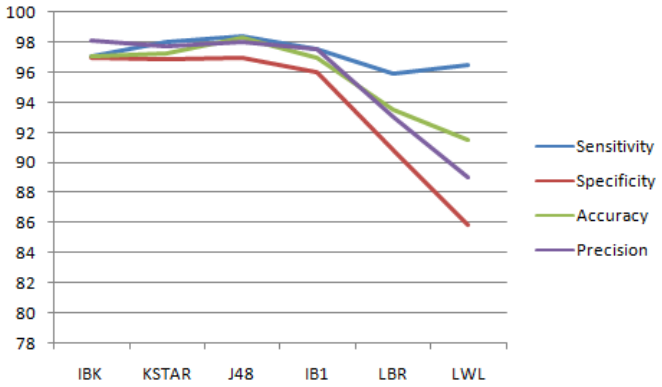


Fig 5.2 – Comparison of different Lazy Classifiers.

Fig 5.3 gives an overview on Meta classifiers. Random committee classifier provides better accuracy measures (97%) for the dataset.

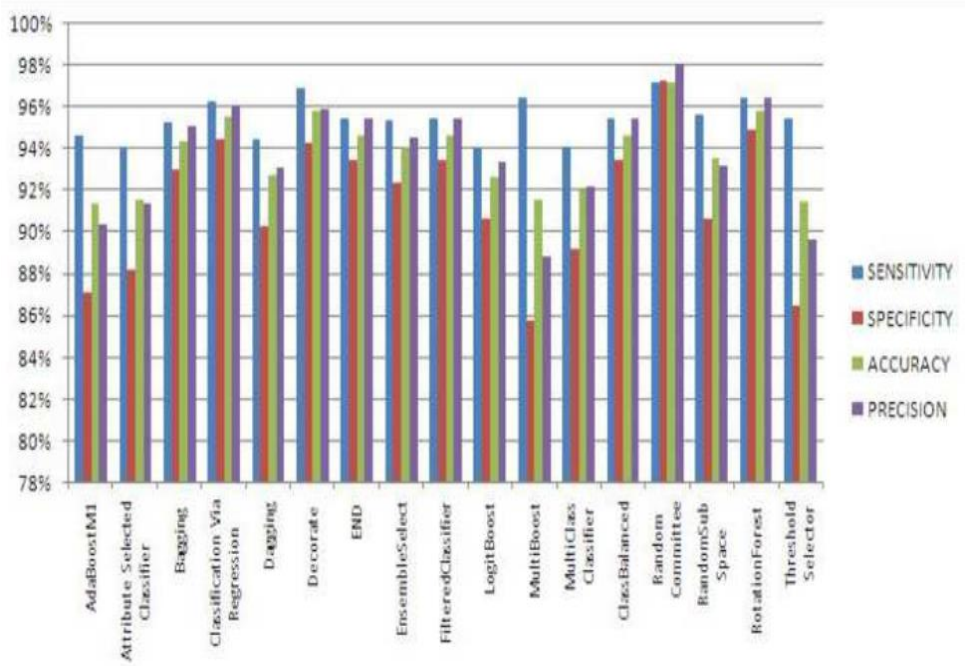


Fig 5.3 – Comparison of different Meta Classifiers.

For instance, J48 leads by sensitivity measure as in Fig 5.4.

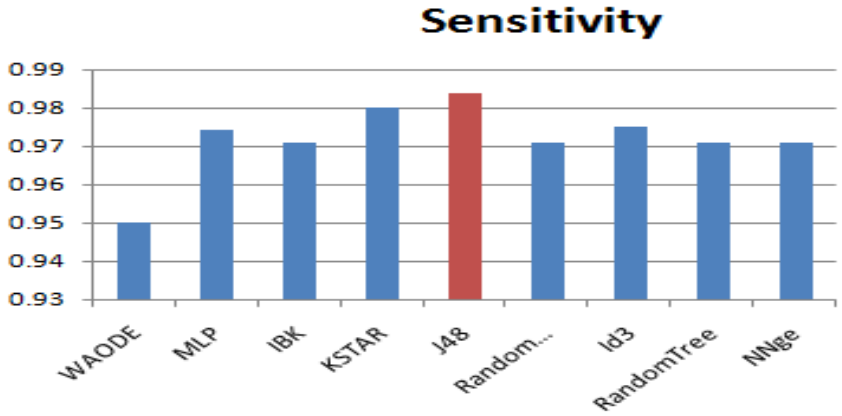


Fig 5.4 – Sensitivity Measure.

ID3 have the best in specificity measure as in Fig 5.5.

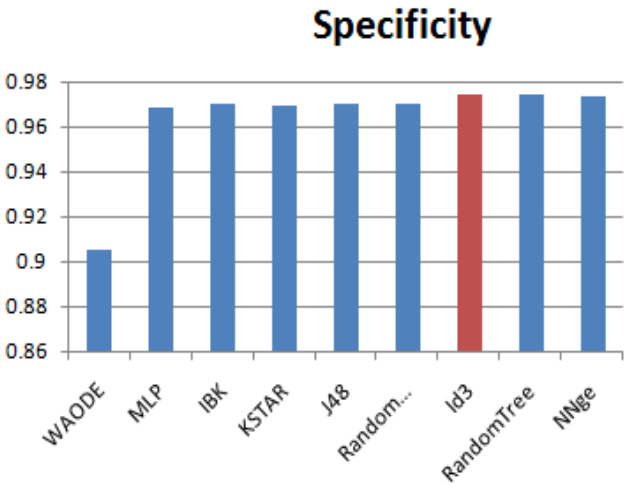


Fig 5.5 – Specificity Measure.

J48 and K Star are the best with accuracy measure as in Fig.5.6.

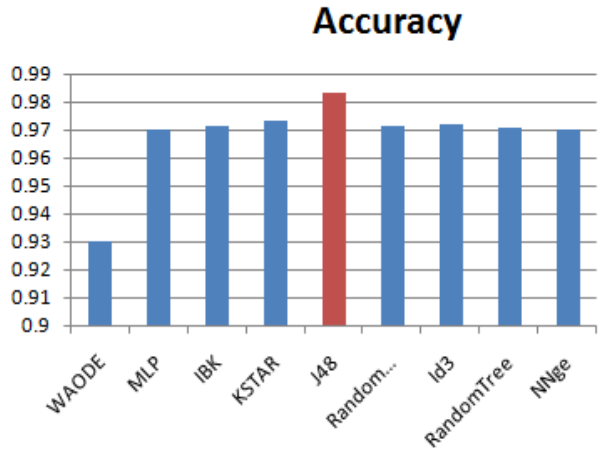


Fig 5.6 – Accuracy Measure

IBK and Random Comitee are the best with precision measure as in Fig. 5.7.

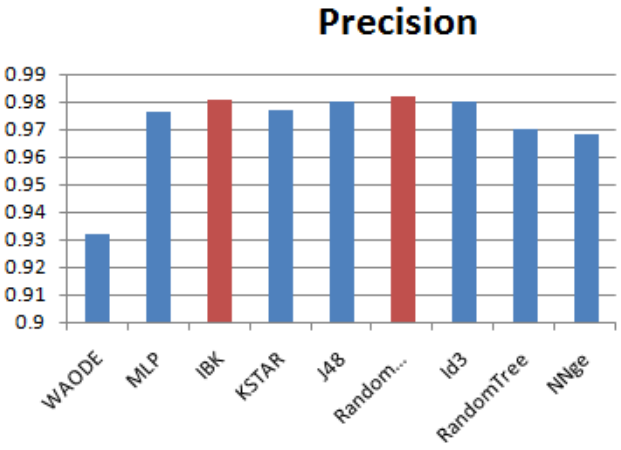


Fig 5.7 – Precision Measure

J48 and Id3 are the best with F-measure as in Fig. 5.8.

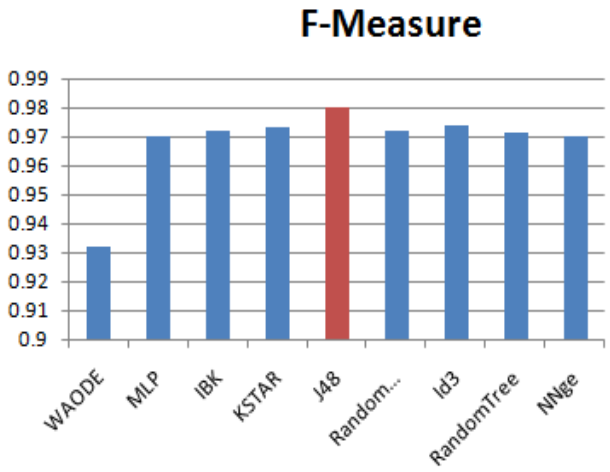


Fig 5.8 – F-Measure

No studies have been in investigation of sudden deaths due to Ischemic Heart disease without major tests like ECG, Stress tests, Cholesterol, Diabetes are found in literature. Literatures includes various tests [25,26,27,28,29,30,31] for the identification of IHD. Through the systematic literature search, three incidence studies on CHD is being identified in the Indian population [32]. As variables for the measurement of social determinants of health depends on various factors like Place of residence, Race/Ethnicity, Gender, Education, Socio-Economic studies[33],further studies with different population is necessary in future

VI. CONCLUSIONS

Based on the results that are obtained from the above, the following conclusions were made:
A set of attributes related to the signs and symptoms and high-risk groups is being proposed which has been tested and which has been proved to be proficient and efficient in recognizing the cases of IHD. The preliminary results of this work suggest that CVD[17][19] is being diagnosed using clinical decision support systems. J48 algorithm helps in appropriate segregation of the negative and the positive cases more approximately than the other methods. Sensitivity and Accuracy are much better than the other methods. It is more essential to have future studies on classifier accuracy and its working proficiency including attribute selection for IHD in developing an electronic protocol for CDSS.
For tracking risk factors population based adult cohorts has been set up:
• Identification of determinants in regardance to the risk factors using cross sectional and prospective designs.
• Associating risk factors with CVDs using prospective design. (e.g., PURE study).
• Case-control study design which helps in rapid assessment of risk factors under adequate funding and supervision.
• Risk factor biology and genomics.

Conflicts of interest
The authors declare that they didnot receive any financial support to conduct this research**.**

Appendix A
Attributes analyzed in DB: signs, symptoms and high-risk groups

Type Signs/symptoms/ high-risk groups	
Physical Examination Parameters	Age
	Gender
	Menopause
	Height
	Weight
	Body mass index
	Waist measure
Co Morbid Features	Systolic Blood Pressure
	Diastolic Blood Pressure
	Diabetes
	Cholesterol
	Thyroid

Personal history	Personal habits(smoking/drinking)
	Family history
	Type A personality/Sleep Disturbance

REFERENCES

[1] Dr.Vaidyanathan and K.Rajeswari, - Artificial Intelligence techniques applied to the development of a clinical decision support system for diagnosing Ischemic Heart Disease., International Journal of Medical Informatics Vol.70.,2012.

[2] KS Reddy and committee - International Heart Protection Summit September 2011**.**

[3] Dr.P.Amirtaraj and K.Rajeswari - Classification of Risk Level for Ischemic Heart Disease in India using Artificial Intelligence., Artificial Intelligence in Medicine., Vol.56.,2011.

[4] Dr.P.Amirtharaj and Dr.Vaidyanathan - Prediction of Risk Score for Heart Disease in India using Machine Intelligence.,. ICMLC 3rd International Conference on MachineLearning and Computing vol7 2011.

[5] Cardiovascular Diseases & its Impact – WHO.,2010.

[6] By Rajeev Gupta & his associates - Burden of Cardiovascular Diseases in India.,2007.

[7] T. Gurumoorthy and K.Rajeswari Modeling Effective Diagnosis of Risk Complications in Type 2 Diabetes – A Predictive model for Indian Situation., Indian J Med Res 132.,2012

[8] Dr.V.Vaithiyanathan and Dr.P.Amirtharaj - A Novel Risk Level Classification of Ischemic Heart Disease using Artificial Neural Network Technique – An Indian Case Study., Artificial Intelligence in Medicine Vol.50.,2012.

[9] K. Rajeswari& V. Vaithiyanathan - Heart disease diagnosis: an efficient decision support system based on fuzzy logic and genetic algorithm., Application to Cardiovascular Diseases_, IEEE Transactions on BioMedical Engineering, Vol. 54.,2010.

[10] Preethi S.J. and K.Rajeswari - Image Enhancement Techniques for Improving the Quality of Colour and Gray scale Medical Images., Artificial Intelligence in Medicine Vol.50.,2009.

[11] K.Rajeswari & V.Vaithiyanathan - Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network., International Journal of Medical Informatics Vol.60.,2011.

[12] Josceli Maria Tenório, Anderson Diniz Hummel, Frederico Molina Cohrs,Vera LuciSdepanian, Ivan Torres Pisa, Heimar de Fátima Marin, ―Artificial intelligence techniques applied to the development of a decision–support system for diagnosing celiac disease‖,International Journal of Medical Informatics Vol. 80.,2011.

[13] Jomini V, Oppliger-Pasquali S, Wietlisbach V, Rodondi N, Jotterand V, Paccaud F ―Contribution of major cardiovascular risk factors to familial premature coronary artery disease: the GENECARD project‖. J Am CollCardiol 2002.

[14] Chih-Lin Chi, W. Nick Street, David A. Katz c, ―A decision support system for cost effective diagnosis‖, Artificial Intelligence in Medicine Vol.50.,2010.

[15] Asuncion A, Newman DJ. ―UCI machine learning repository‖. Accessed at MLRepository., July 16, 2012.

[16] Liu GP, Wang YQ, Dong Y, et al: Development and evaluation of Scale for heart system Inquiry of TCM. Journal of Chinese Integrative Medicine,Vol.7PP:1..2009.

[17] International Society of Cardiology and the Joint Subject Group on standardization of clinical naming in World Health Organization: Naming and diagnosis criteria of ischemicheart disease. Circulation Vol. 59 PP:3.,2012.44

[18] Vamadevan S. Ajay &DorairajPrabhakaran, _Coronary heart disease in Indians: Implications of the INTERHEART study_, Indian J Med Res 132,pp 561-566. , November 2010

[19] Reddy KS, Shah B, Varghese C, Ramadoss A. Responding to the threat of chronic diseases in India. Lancet Vol.366.,PP1744-9.,2012.

[20] Gupta R, Joshi P, Mohan V, Reddy KS, Yusuf S. Epidemiology and causation of coronary heart disease and stroke in India. Heart;Vol: 94 :PP: 16-26.,2008.

[21] "Obese adolescents have heart damage." obeseadolescents-heart PHYSorg.com. (accessed on July 15, 2012).

[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten,The WEKA data mining software: an update, SIGKDD Explorations 11 (1)(2009).

[23] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques,2nd ed., Morgan Kaufmann, San Francisco, 2005.

[24] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification of Regression Trees.Wadsworth, 1984.

[25] J. R. Quinlan. Induction of decision trees. Machine Learning, 1:81–106,1986.

[26] J. R. Quinlan. C4.5: Programs for Machine Learnin. Morgan Kaufman,1993.45

[27] Ajay VS, Prabhakaran D, Jeemon P, Thankappan KR, Mohan V,RamakrishnanL,Prevalence and determinants of diabetes mellitus in the Indian Industrial population. Diabetic Med ; Vol.25 .,PP1187-94.,2008.

[28] García-Nietoa,E.Alba a,,L.Jourdanb,E.Talbi b, _Sensitivity and specificity based multiobjective approach for feature selection: Application to cancer diagnosis_, Information Processing Letters Vol:109 PP:887–896.,2009.

[29] Newman, D.J., Hettich, S., Blake, C.L., Merz, C.J., UCI repository of machine learning databases. Department of Information and Computer Science, University California Irvine., 1998.
[30] E. Massad, A teoriabayesiana no diagnósticomédico, in: E Massad, RX Menezes, PSP Silveira, NRS. Ortega (Eds.),
[31] Métodosquantitativosemmedicina, Barueri (SP), Manole, pp. 189–205.,2004
[32] P.K. Anooj, _Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules_, Journal of King Saud University –Computer and Information Sciences Vol. 24,PP 27–40.,2012.
[33] European ST-T Database Directory. Pisa, Italy: S.T.A.R., 1991
[34] Markos G. Tsipouras, Costas Voglis, and Dimitrios I. Fotiadis, _A Framework for Fuzzy Expert System Creation—Application to Cardiovascular Diseases_, IEEE Transactions on BioMedical Engineering, Vol. 54, No. 11, November 2007.
[35] Ali Gharaviri, Member, Mohammad Teshnehlab and H. A. Moghaddam,_Ischemia Detection via ECG Using ANFIS_, 978-1-4244-1815-2/08/$25.00©2008 IEEE.46
[36] Minas A. Karaolis, Member, Joseph A.Moutiris, DemetraHadjipanayi, and Constantinos,S. Pattichis, _Assessment of the Risk Factors of Coronary Heart Events Based on Data MiningWith Decision Trees_, IEEE Transactions on Information Technology in BioMedicine, Vol 14, No 3,May 2010.
[37] D Frenkel, J Nadal, _Ischemic Episode Detection using an Artificial Neural Network Trained with Isolated ST-T Segments_, 0276-6547/99 $10.00 0 1999 IEEE.
[38] MinghaoPiao, HeonGyu Lee, Guo Yong Sohn, GoucholPok, KeunHo Ryo,_EmergingPatterns based Methodology for Prediction of Patients with Myocardial Ischemia_, 978-0- 7695-3735-1/09 $25.00 © 2009 IEEE.
[39] National Cardiovascular Disease Database, Sticker No: SE / 04 / 233208,IC Health, Supported by Ministry of Health & Family Welfare, Government of India and World Health Organization.
[40] Panniyammakal Jeemon& K.S. Reddy, _Social determinants of cardiovascular disease outcomes in Indians_, Indian J Med Res 132, pp 617-622. , November 2010.

